

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
17 October 2002 (17.10.2002)

PCT

(10) International Publication Number
WO 02/082716 A1

(51) International Patent Classification⁷: **H04L 9/00**,
H04K 1/00, G06F 19/00

(21) International Application Number: PCT/US01/10886

(22) International Filing Date: 2 April 2001 (02.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US):
GEOTRUST, INC. [US/US]; Suite 1650, 700 NE
Mulnomah Street, Portland, OR 97232 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **MCLEOD, Scott**,
C. [US/US]; 24 Carriage Drive, Chelmsford, MA 01824
(US). **NORMAN, Peter, D.** [US/US]; 56 Palmer Street,

Arlington, MA 02174 (US). **WILLOUGHBY, Kevin**
[US/US]; 10 Church Street, Framingham, MA 07102
(US). **ROSENBERG, Jonathan, B.** [US/US]; 11 Seton
Hall Road, Auburndale, MA 02466 (US). **COULTHARD**,
Christopher, M. [GB/US]; 88 Park Ave., #402, Arlington,
MA 02476 (US).

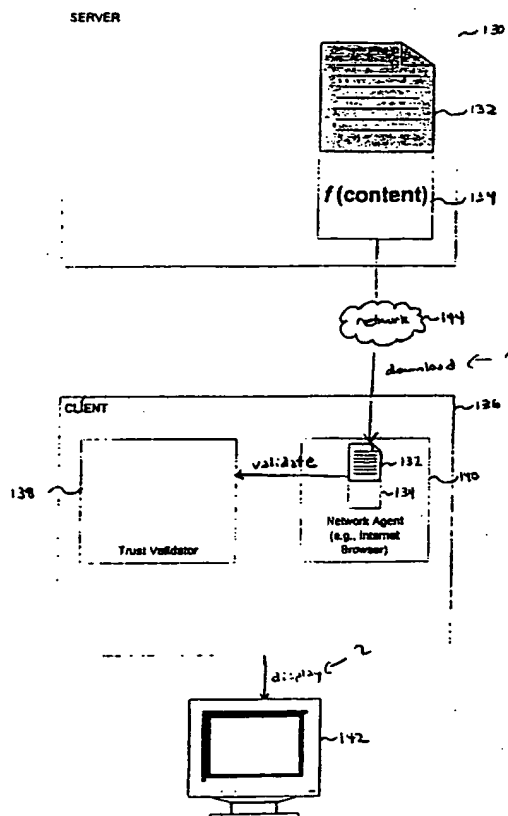
(74) Agent: **CANNAVALE, Stephen**; Goodwin Procter LLP, 7
Becker Farm Road, Roseland, NJ 07068 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian

[Continued on next page]

(54) Title: VALIDATING CONTENT



(57) Abstract: A method of processing content includes storing verification information corresponding to certified content at a first computer (100) and receiving a verification request corresponding to content (110) from a second computer (130). The method also includes determining a verification information for the content corresponding to the verification request and comparing the determined verification information with the stored verification information.

WO 02/082716 A1

BEST AVAILABLE COPY



patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— with international search report

VALIDATING CONTENT

REFERENCE TO RELATED APPLICATIONS

5 This application claims priority from U.S. provisional application Serial No. 60/153,901, filed September 14, 1999. This application is also a continuation-in-part of pending U.S. application Serial No. 09/248,370, entitled "Content Certification", filed on February 8, 1999; pending U.S. application Serial No. 09/312,751, entitled "Dynamic Content Certification, filed on May 13, 1999; and pending U.S. application Serial No. 10 09/311,607, entitled "Defining Content Zones, filed on May 13, 1999. All of these applications are incorporated by reference in their entirety herein.

BACKGROUND

Many recent news stories have brought the credibility of information on the Internet into question. These stories have ranged from pranks such as M.I.T. students altering their school's home-page on April Fools Day to more troublesome security breaches such as the 15 tampering of the United States Navy's web-site during the Kosovo crisis.

Even when not subjected to hacker mischief, many web-site providers have difficulty maintaining control over the content offered by their site. For example, some employers give employees fairly free reign in posting content. Unfortunately, without sufficient review, this 20 often results in postings of dubious quality. This can injure the business or organization providing the site and the overall reputation of the reliability of Internet information.

SUMMARY

In general, the invention features a method of processing content received from a networked computer in response to a browser request for content. The method includes the 25 steps of receiving certification information associated with content received by the browser; determining a certification status for content based on the received certification information; and displaying at least one indication of the determined certification status of the content.

In preferred embodiments, the indication may include a persistent indication displayed with the content. The indication may include a taskbar button. The indication may 30 include a tray icon. Displaying at least one indication may include processing the content to

include one or more indications. Processing the content comprises altering visual representation of the content.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1 is a screenshot of a browser's display of an Internet page.

FIGS. 2-8 are screenshots of different persistent displays that notify a user whether content is certified.

FIGS. 9, 11, 13, and 15 are diagrams of systems for validating content certification.

FIGS. 10, 12, and 16 are flow-charts of processes for validating content certification.

FIG. 14 is a diagram of a manifest of web-page contents.

FIG. 17 is a diagram of a certification server and a validation server.

DETAILED DESCRIPTION

Introduction

FIG. 1 shows a web-page 100 presented by an Internet browser. A web-page can feature an assembly of different content such as graphics, text, audio, video, and even software instructions such as Java Applets and ActiveX controls. As shown in FIG. 1, a user viewing the page 100 often must trust that the content-provider stands behind the contents and/or that the contents have not been tampered with. Sometimes this trust is misplaced. For example, someone may have posted the content at the business' web-site without appropriate approval (e.g., undergoing a certification process). Alternatively, some intermediate network node may have intercepted content as it traveled across the Internet and replaced selected portions.

Co-pending applications "Content Certification", "Dynamic Content Certification", and "Defining Content Zones" describe techniques that enable a content provider to certify content. These applications also describe techniques for validating certification of downloaded content. Such validation can include determining content is not certified, determining content was altered after certification, determining certification has expired, and/or determining certification has been revoked. Such validation can also include

determining and authenticating the identities of entities claiming to have certified the content. As shown in FIGS. 2-8, these techniques have been embodied in a software program that can use graphical indicators, sound, and other notification techniques to notify a user whether downloaded content is certified content.

5

Display of Certification Status

A number of different mechanisms can notify users of whether downloaded content is certified content. For example, FIGS. 2 and 3 show a Microsoft[®] Windows 95 taskbar button 104 and tray icon 106 that change appearances based on an attempt to validate certification of content displayed in an active browser window. For example, the controls 104, 106 may notify a user of the certification status (e.g., certified, uncertified, expired, revoked, etc.) of content using text, graphics, color, and other display attributes. The appearance of the controls 104, 106 may vary in different ways for different certification statuses. For example, content that was never certified may cause the tray icon to display a bright red skull and cross bones to alert a user, while content having revoked certification may cause the tray icon to turn orange. The unobtrusive placement of the controls 104, 106 provides real-time, continual, notification of content certification without interfering with a user's normal browser interaction.

FIGS. 4-7 show a number of other user notification techniques. For example, FIG. 4 shows a window 108 that displays a map 110 of content displayed by a browser. The map 110 may include a logo (not shown) of the site offering the content. The different appearances of map regions indicate the certification status of content. For example, red portions may indicate uncertified regions of a page, while white portions may indicate certified regions. The window enables a user to quickly identify potentially uncertified content.

FIG. 5 shows a window 112 that displays a tree of web-page contents 114-120. Each node in the tree can correspond to a different content (e.g., a node for a page's HTML and nodes for different GIF (Graphics Interchange Format) pictures referred to by the page). Again, different display attributes of tree nodes reflect the certification status of content. For example, shaded node 116 indicates that the picture for "Digests of Patent Opinions Federal Circuit" has not been certified. The map of FIG. 4 and the tree of FIG. 5 can provide a user

with a visual description of content certification, without altering the browser's display of the page or otherwise altering the browser's functions.

Other techniques, however, use browser-provided functions to provide an indication of the certification status of content. For example, as shown in FIG. 6, a browser may be
5 dynamically programmed to display the certification status of content on a page as a user brushes the content with a cursor. For browsers not offering this capability, this feature may be offered by continuously determining cursor placement and displaying a window near the content. Alternatively, the window may only be displayed when a user selects content, for example, by clicking a mouse button on the content.

10 As shown in FIG. 7, software can also directly alter the display of contents after determining the certification of different portions. For example, as shown, the software can black-out 114 uncertified content, and/or alter the display of content 116 having expired certification. Depending on the browser, this may require writing a downloaded page to a temporary file, modifying the temporary file, and reloading the modified temporary file into
15 the browser.

The embodiments described above can also provide more detailed information about the certification of content. For example, by selecting the system taskbar button 104 in FIGS. 2 or 3, a dialog, as shown in FIG. 8, can display detailed information about content. The detailed information can include the certifying entity 124, a graphic for the entity (e.g., a
20 business trademark), the trustworthiness of the page or content 125, the URL (Universal Resource Locator) or URI (Universal Resource Indicator) of the content 127, the range of dates the certification is valid 128, and a "digital fingerprint" of the content 129. The dialog may also display other information (not shown) such as the site certificate of the web-site providing the page and potentially a text description of the "Trust Policy" used by the site to
25 certify content (e.g., "Factpoint, Inc. uses a five person review board to certify content prior to posting").

Any of the visual techniques described above can be combined and/or used in conjunction with non-visual techniques such as audio messages (e.g., "The picture of Abe Lincoln is untrustworthy"). Additionally, while the above description described individual
30 pages, the same techniques work equally well with framed browser displays that display two or more pages simultaneously.

Underlying the displays shown in FIGS. 2-8 are certification procedures that enable providers to certify posted content and validation procedures that enable users to validate the certification of received content. Again, many of the techniques have been described in the co-pending U.S. patent applications.

5

The Trust Validator

FIG. 9 shows a client 136 browser 140 downloading information (i.e., page 132) from a URL (Universal Resource Locator) 132 over a network 144. The client 136 can present the downloaded content on a user's monitor 142, speaker, etc. As shown, the client 136 includes
10 "trust validator" software 138 that validates certification of downloaded content. The validator 138 may operate as a background process that monitors content received by the browser 140, for example, via calls to or from the browser API (application programming interface). Alternatively, validator 138 functions may be directly integrated into the browser 140.

15 The validator 138 can validate content certification using certification information associated with the content. For example, the validator 138 can compare certification information determined for the content determined prior to transmission to the client with certification information determined after transmission.

In more detail, a certification process produces certification information 134 based on
20 the certified content(s). Typically, this information 134 is produced using a "one-way" function. For example, a hashing function may use all or some portion of the ASCII characters in HTML (HyperText Markup Language) commands that define a page to produce a set of output bytes. Given the same input, the hashing function produces the same output. A popular hashing functions known as MD5 and SHA can produce relatively small output for
25 large pages.

The certification information 134 derived from the content may be included in the content itself, for example, as data, for example, as signature and/or manifest elements of an XML (Extensible Markup Language) page or as an HTML "Meta" element. When the certification information 134 is included in the content, it must be removed before re-
30 determining the certification information.

Alternatively, the information 134 may be included in the header of an HTTP (HyperText Transfer Protocol) message sent by the server 130. In yet another implementation, the trust validator 138 may independently request certification information 134 for the downloaded content. For example, the site may provide a file (e.g.,
5 "factpoint.txt") at a predefined location (e.g., "www.url.com/factpoint.txt") that lists where certification information 134 for site content can be found. The file may refer to other sites when the content has been copied.

FIG. 10 shows a process 138 the trust validator can use to validate certification of downloaded content. First, the trust validator obtains 150 the downloaded content (e.g., a
10 page or portion of a page) and the certification information associated with the content. The trust validator 138 can obtain this information from the browser 140 or can establish an independent connection with the server 130. The trust validator 138 can independently determine certification information using 152 the one-way function on the received content. By comparing 154 the received certification information and the independently determined
15 certification information, the validator 138 can determine 154 whether the page 132 has been altered since certification and notify a user of such a change. The trust validator may also notify a web-site administrator if certification validation fails so the administrator can investigate uncertified content offered by the site.

FIG. 11 shows a scheme that cannot only detect tampering, but that can also identify
20 and authenticate the entity or entities certifying content. This scheme features certification information that includes a hash digitally signed by one or more certifying entities. A digital signature 160, much like a handwritten signature on a piece of paper, provides a degree of certainty that a particular entity signed the content in question.

One digital signature scheme uses a private encryption key known only to the signer
25 and a public encryption key that may be freely distributed. Information encrypted with the private key can only be unencrypted with the public key. Thus, an entity certifying content can encrypt a hash of the content with their private key. Only the public key associated with the entity can properly decrypt the hash. For example, a hash of content may be encrypted using a private key assigned to a web-site and decrypted using a public key included in the
30 site's certificate. A wide variety of other digital signature schemes may be used such as an exchange of a single encryption key or the use of physical devices such as smart cards.

In the system of FIG. 11, information needed to validate a digital signature may be included with the certification information. The information may include an X.509 certificate for each entity signing the hash. For example, an X.509 certificate may include the public key needed to decrypt the hash of the page 132, a description of the entity holding the private key, and the digital signature of some authority such as VeriSign7 testifying to the truth of the information in the certificate (i.e., that the entity claiming to have signed the hash is actually the claimed entity). In another embodiment, the information needed to validate a digital signature (or a reference to this information) may be provided by one or more DSig (Digital Signature Users Group) digital signature blocks.

As shown in FIG. 12, after receiving the certification information (e.g., digital signature and certificates), the trust validator 138 can use the public key included in the certificate to extract the hash included in the digital signature. The trust validator 138 can also follow the chain of authority 162, for example, by asking VeriSign7 if the public key received is really the public key of the entity claiming to have signed the hash. The trust validator can include information about the chain of authority in a display such as the dialog shown in FIG. 8. After extracting the hash from the certification information, the trust validator 138 can conclude the page was altered or was never certified to begin with and can notify a user using the techniques described above.

If the certification information includes a digitally signed hash, the certification information may be transmitted over an insecure connection. If, however, the certification information only includes a hash, a secure connection such a secure sockets layer (SSL) connection may be preferred.

As shown in FIG. 13, instead of a single digital signature or hash, certification information may include a manifest 170 for content included in a page. The manifest 170 itself may be hashed and digitally signed. As shown in FIG. 14, the manifest 170 can include the hash values of different page 130 content. For example, the manifest 170 shown includes a different hash value for each picture displayed on the page. The trust validator 138 can use this information to validate each portion of a page individually. The validator 138 can also use criteria to produce an overall estimation of page certification. This criteria may be provided by rules included in the manifest 170 (e.g., defining valid content collections), logic hard-coded into the validator, and/or as logic provided by user-supplied code (e.g., a Java

script). By default, the validator 138 can describe the page as having the lowest certification status of any content in the page. For example, if any content on the page has expired, the page as a whole is deemed expired. The validator 138 may use similar logic for frames. That is, the overall certification status of a display is determined by the worst certification status of any content in any displayed frame.

In some implementations, the trust validator 138 can alert a user to revocation, expiration, and other certification statuses of downloaded content. FIG. 15 shows a server 130 that includes a database table 182 describing available content 132. The table 182 can include an expiration date for certification, a blanket revocation of certification, and other information. Upon receiving content, the trust validator 138 can transmit a validation request to validation software 180 on the server 130. The validation software 132 can access the table 182 to verify the content was certified and determine whether the content has expired or has been revoked. The validation software 132 can transmit the results back to the trust validator 138.

Though information in the table 182 may be included in the certification information received by the client, the table 182 enables an administrator to centrally alter certification information. The server table 182 can also be used to provide content "versioning". For example, a web-site may certify a more recent version of information for a URL. Validation software can look for valid versions of a URL when a client attempts to validate expired or revoked content.

FIG. 16 describes this validation process in greater detail. After receiving the content and its corresponding certification information 200 and independently determining the certification 204 for the content, the validator 138 can preliminarily determine if the content is certified without accessing the server 130. For additional validation, the validator 138 can also transmit 206 certification information (e.g., the hash) to the server validation software for look-up in the server table 182. The server table 182 can not only verify that the content has not expired or been revoked, the server table 182 can also identify more recent content that replaces the content the user downloaded (e.g., the URL for the hash submitted has another table entry that has not been revoked). The trust validator can then establish a connection to download the valid version for display in the browser.

FIG. 17 shows a secure architecture that distributes server certification and validation functions between a certification server 218 and a validation server 232. The certification server 218 includes certification software 220 that certifies submitted content 214 as described in the co-pending patent applications. The certification server 218 also adds table
5 182 entries as content is certified. An administration tool 216 can manage information stored in the table, for example, to specify an expiration date, delete certification, or revoke certification for content.

The certification software 220 may certify a single piece of content or a collection of web-pages using a certification "spider." As described in the co-pending U.S. Applications,
10 certification may be performed for fixed or dynamically constructed content. After certification, the certification server can place certified content on the validation server for distribution.

The validation server 232 includes validation software 228 that accesses the certification server 220 table 182 in response to client validation requests. The validation
15 server 232 may maintain a cache of validation data to reduce the time spent serving client requests.

The techniques described here are not limited to any particular hardware or software configuration; they may find applicability in any computing or processing environment. The techniques may be implemented in hardware or software, or a combination of the two.
20 Preferably, the techniques are implemented in computer programs executing on programmable computers that each include a processor, a storage medium readable by the processor (including volatile and non-volatile memory and/or storage elements), at least one input device, and one or more output devices. Program code is applied to data entered using the input device to perform the functions described and to generate output information. The
25 output information is applied to one or more output devices.

Each program is preferably implemented in a high level procedural or object oriented programming language to communicate with a computer system. However, the programs can be implemented in assembly or machine language, if desired. In any case, the language may be a compiled or interpreted language.

30 Each such computer program is preferable stored on a storage medium or device (e.g., CD-ROM, hard disk or magnetic diskette) that is readable by a general or special purpose

programmable computer for configuring and operating the computer when the storage medium or device is read by the computer to perform the procedures described in this document. The system may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so

- 5 configured causes a computer to operate in a specific and predefined manner.

Other embodiments are within the scope of the following claims.

What is claimed is:

- 1 1. A method of processing content received from a networked computer in response
2 to a browser request for content, the method comprising:
3 receiving certification information associated with content received by the browser;
4 determining a certification status for content based on the received certification
5 information; and
6 displaying at least one indication of the determined certification status of the content.
- 1 2. The method of claim 1, wherein the indication comprises a persistent indication
2 displayed with the content.
- 1 3. The method of claim 1, wherein the indication comprises a taskbar button.
- 1 4. The method of claim 1, wherein the indication comprises a tray icon.
- 1 5. The method of claim 1, wherein displaying at least one indication comprises
2 processing the content to include one or more indications.
- 1 6. The method of claim 5, wherein processing the content comprises altering visual
2 representation of the content.

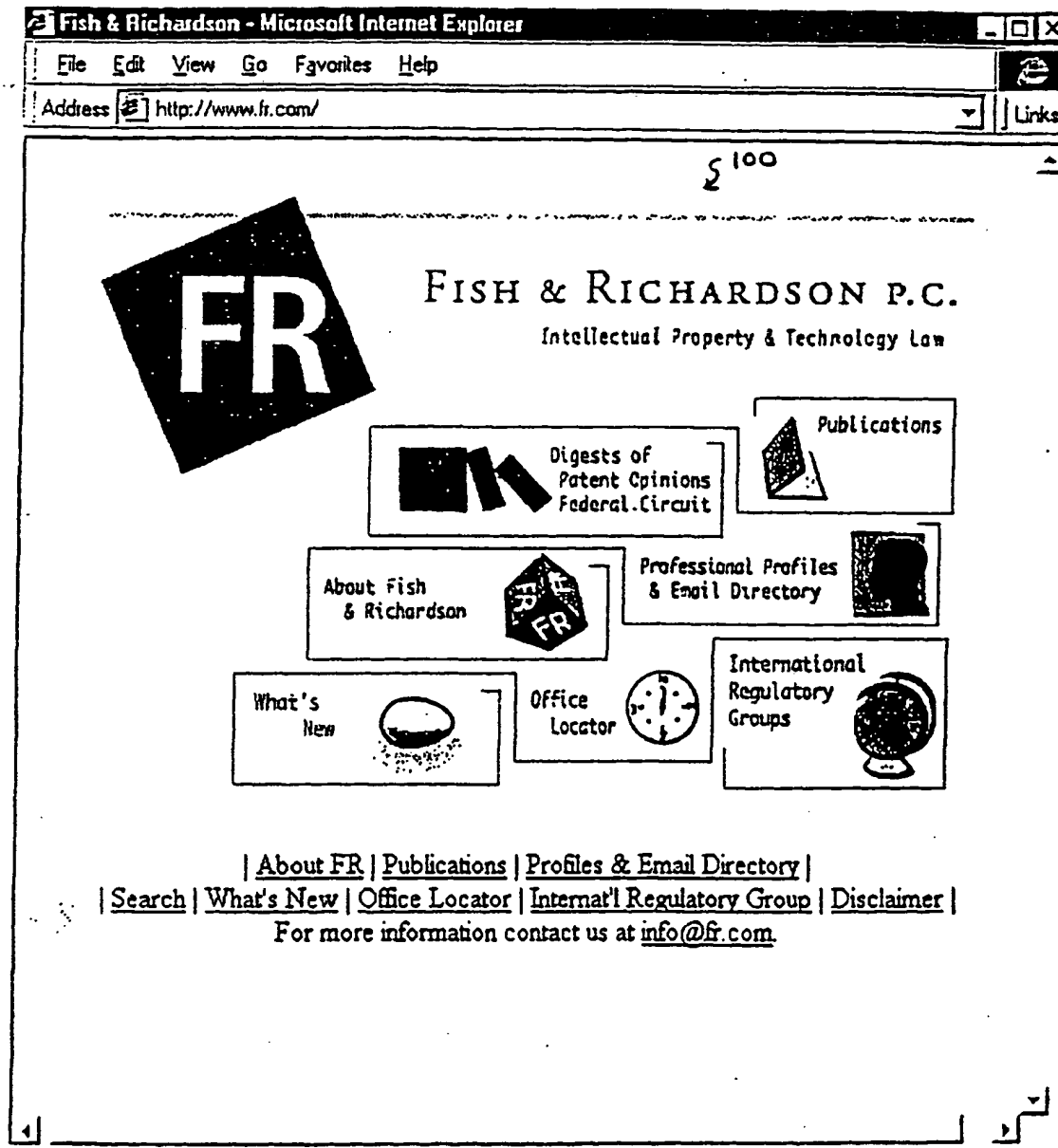


FIG. 1 (PRIOR ART)

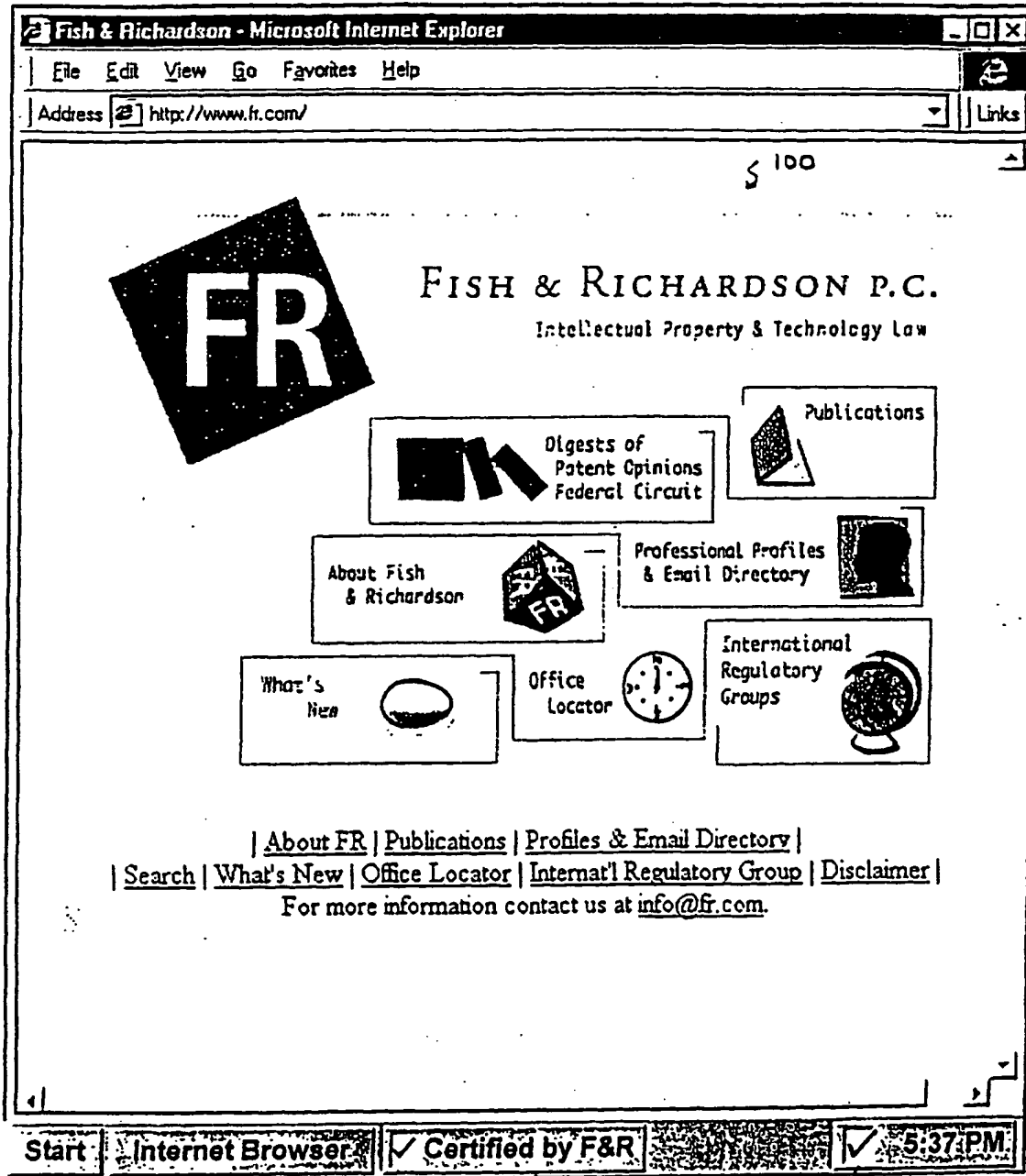
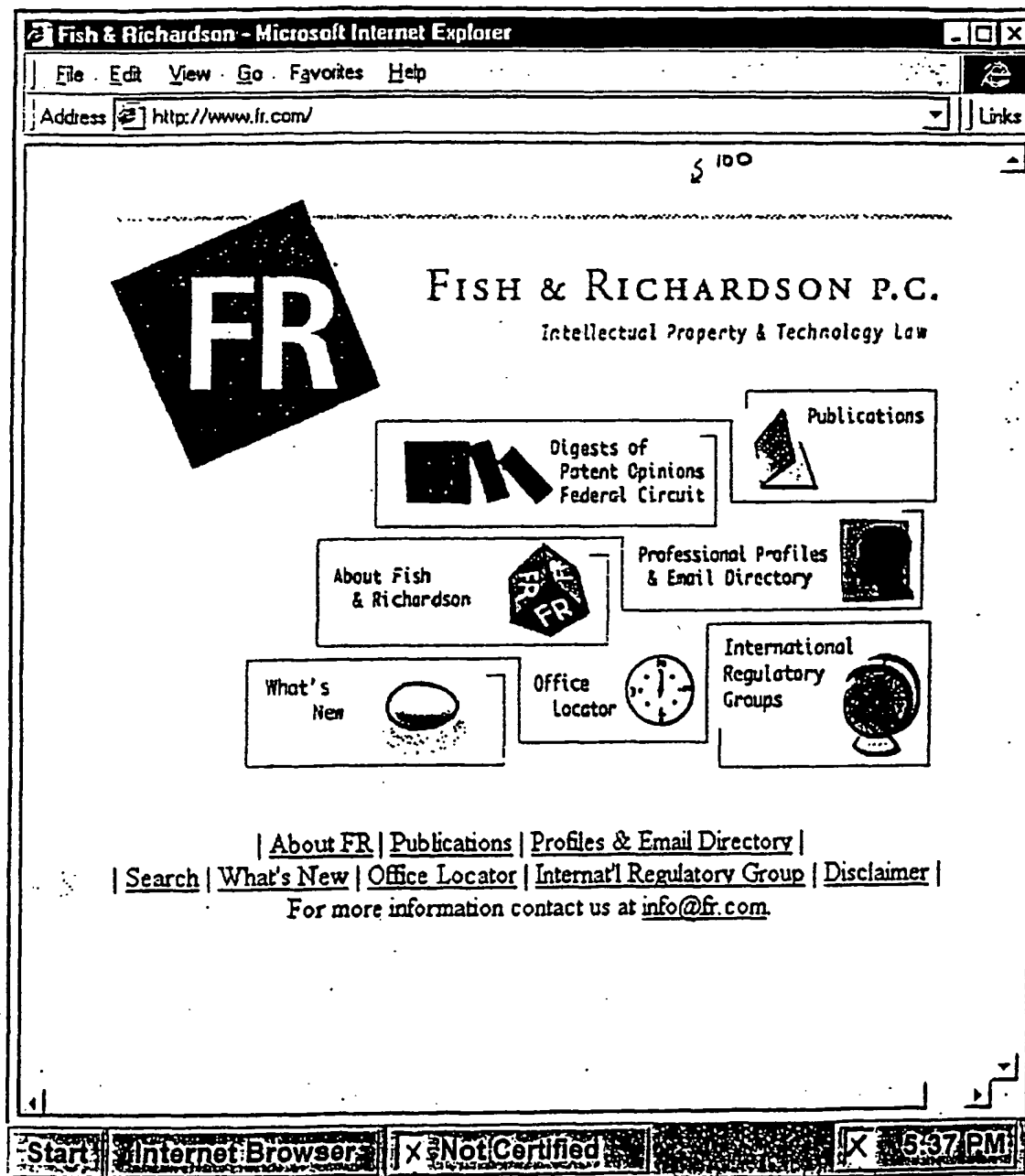


FIG. 2



104

106

FIG. 3

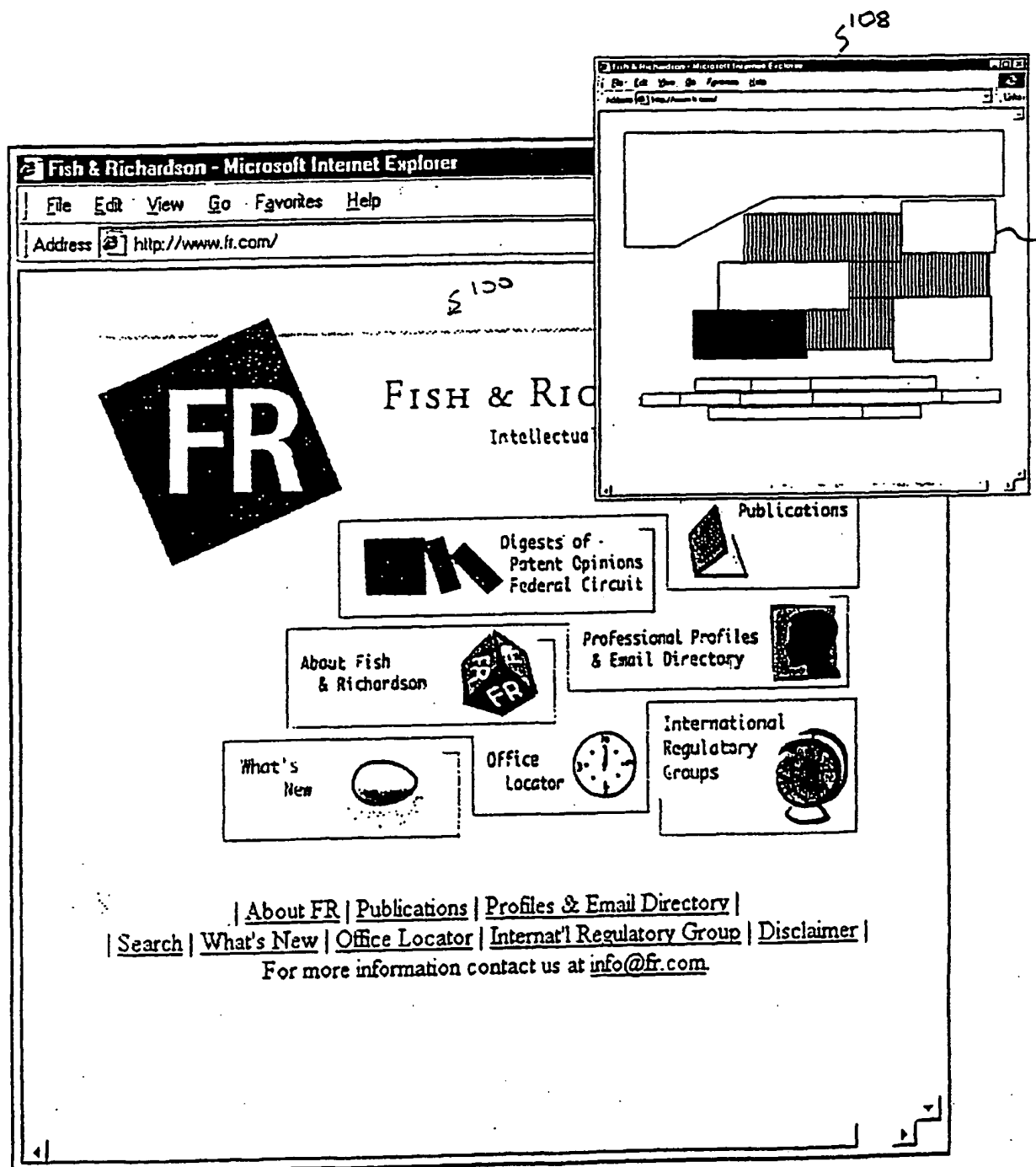


FIG. 4

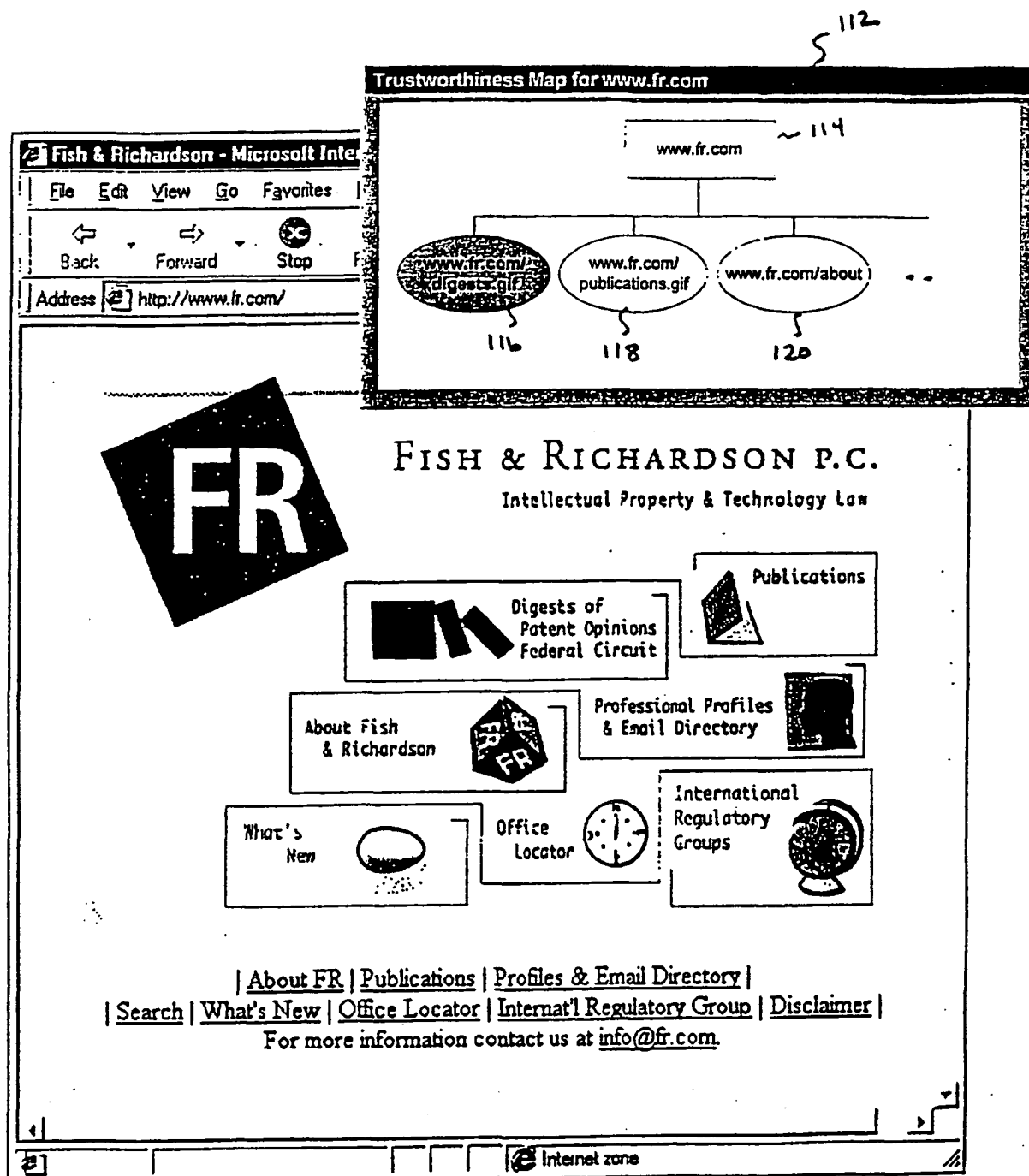


FIG. 5

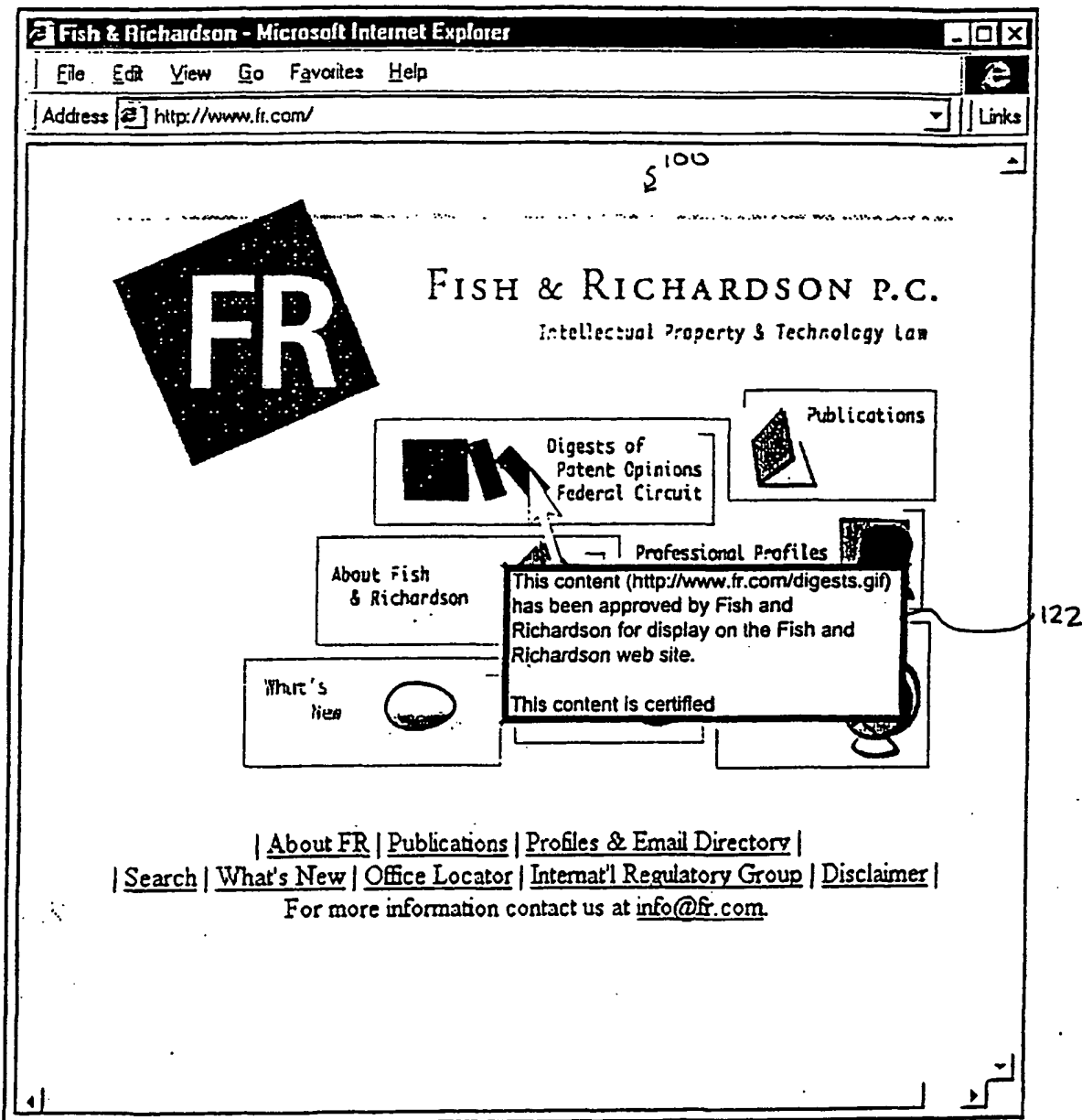


FIG. 6

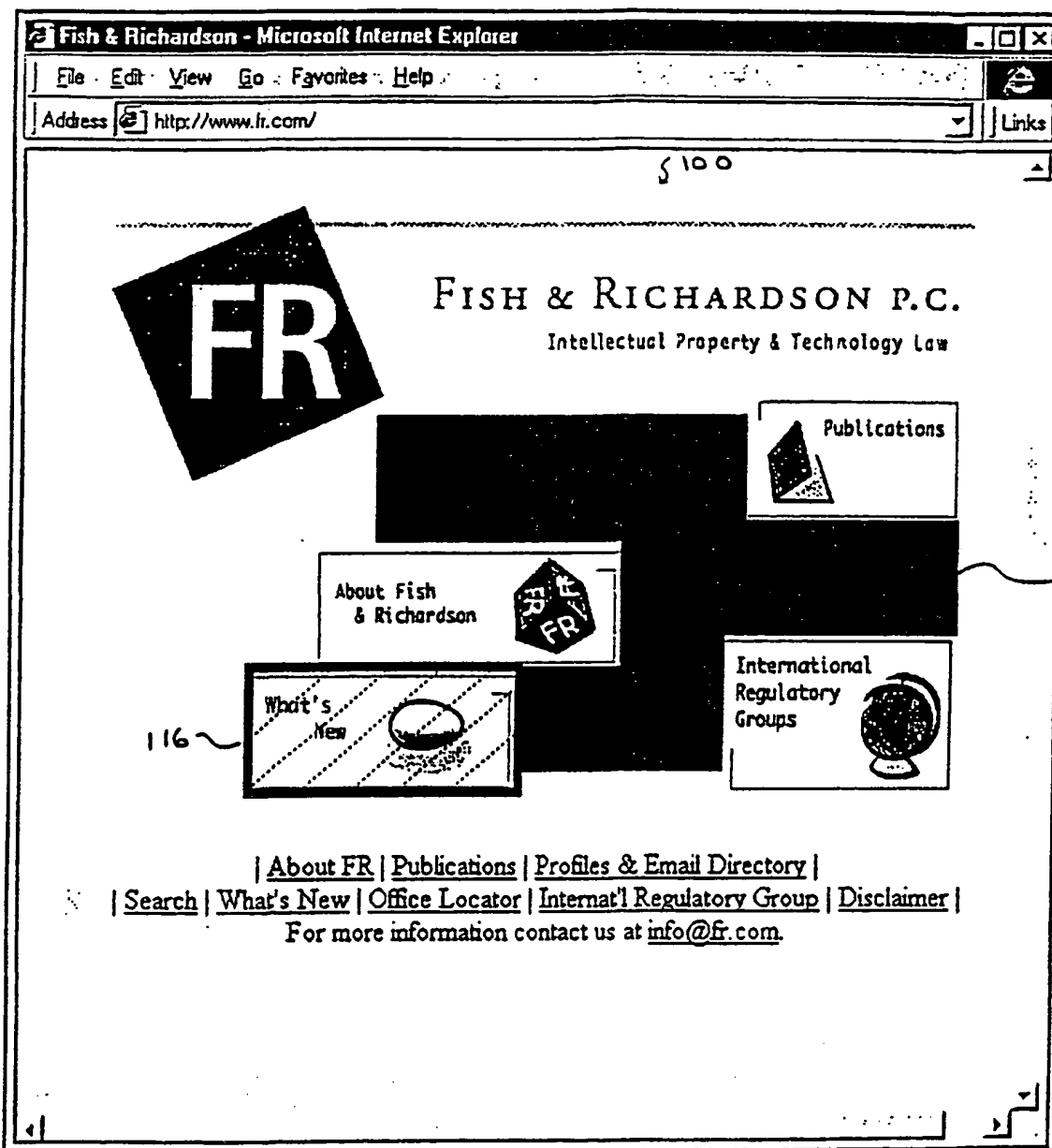


FIG. 7

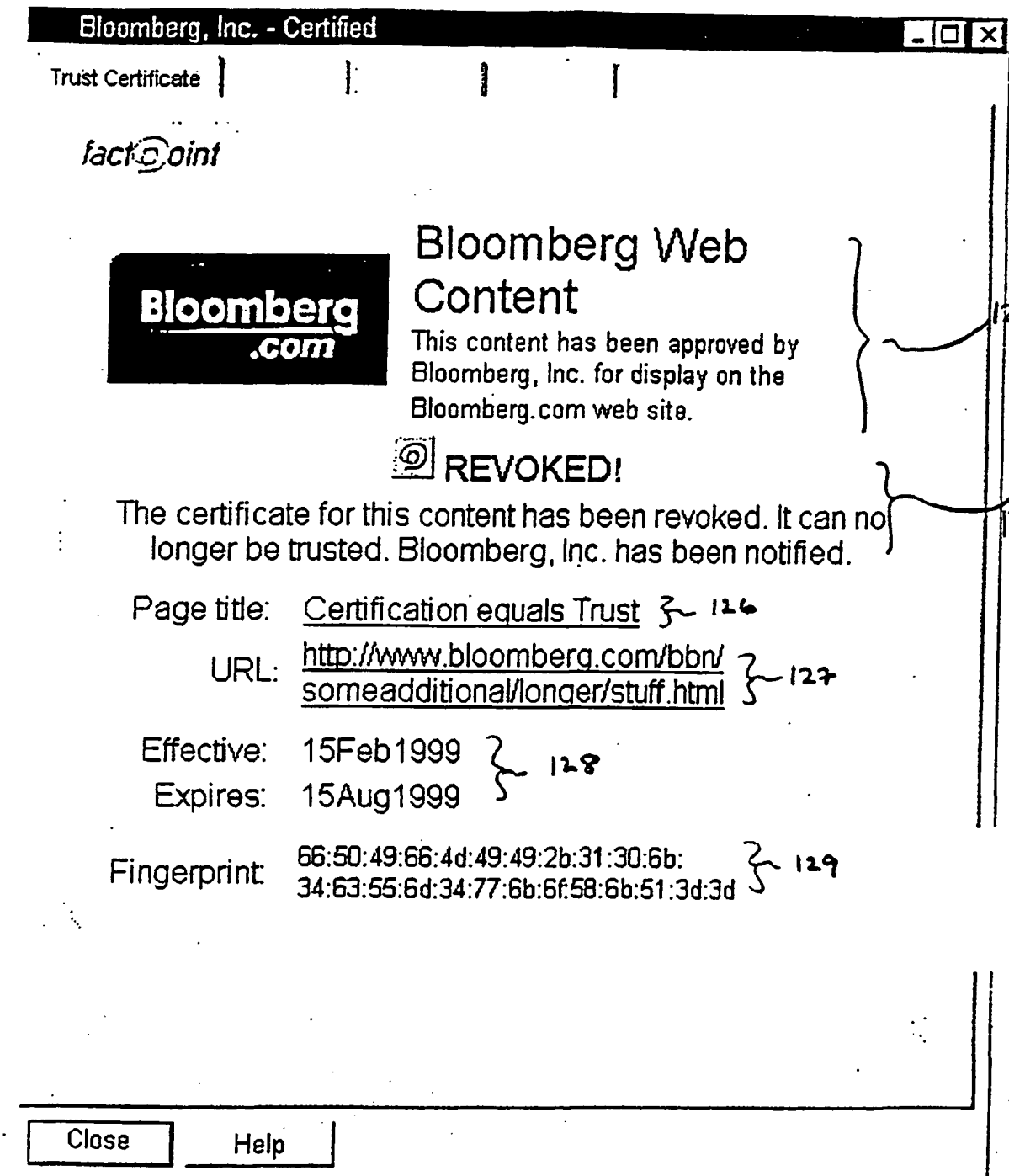


FIG. 8

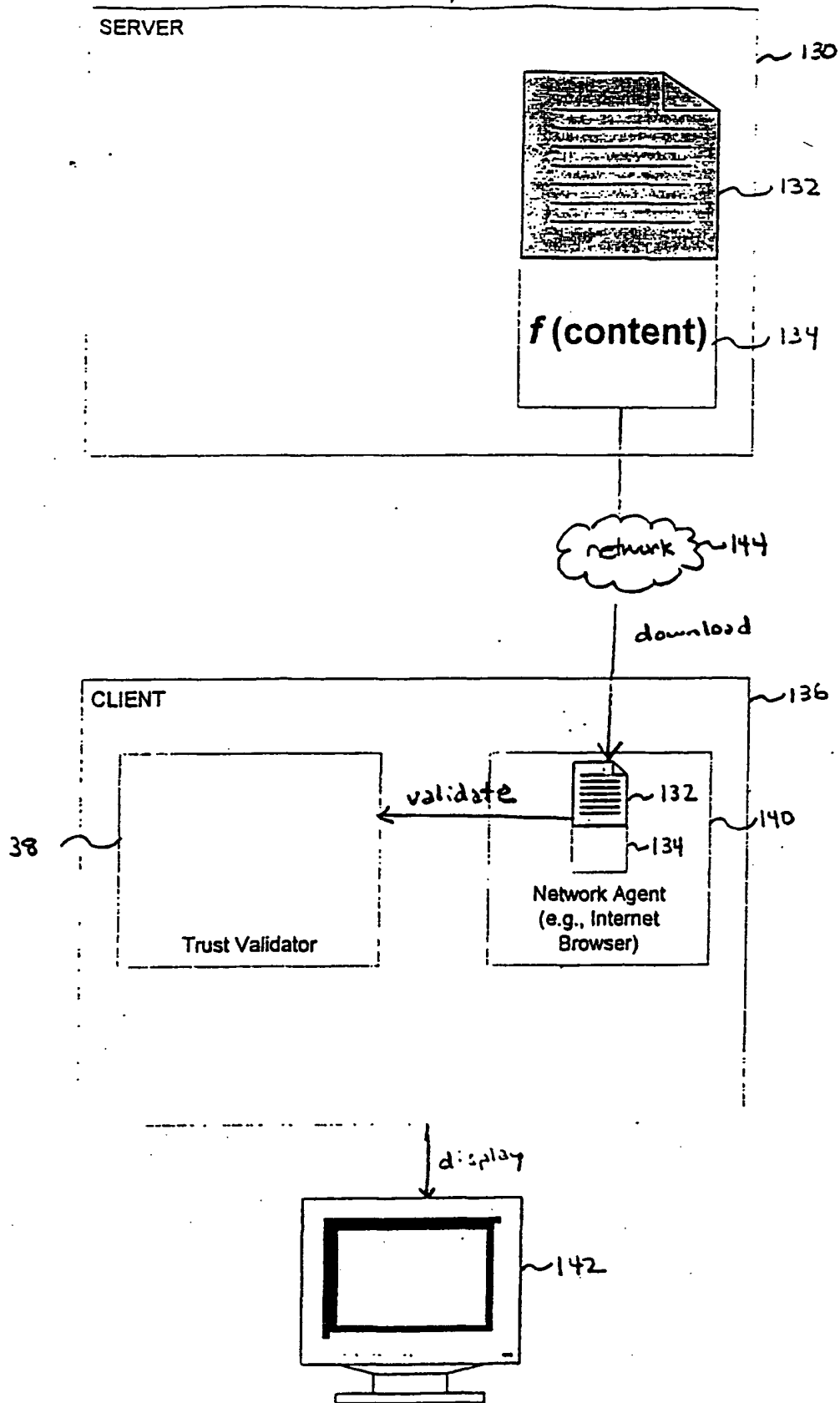


FIG 9

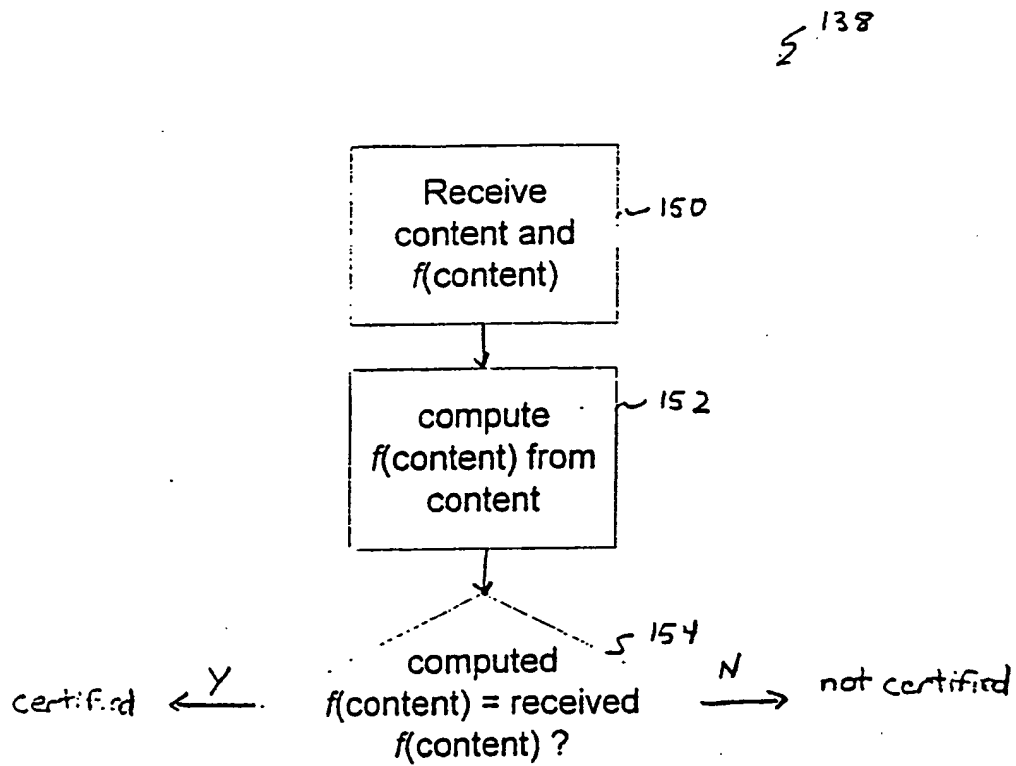


FIG. 10

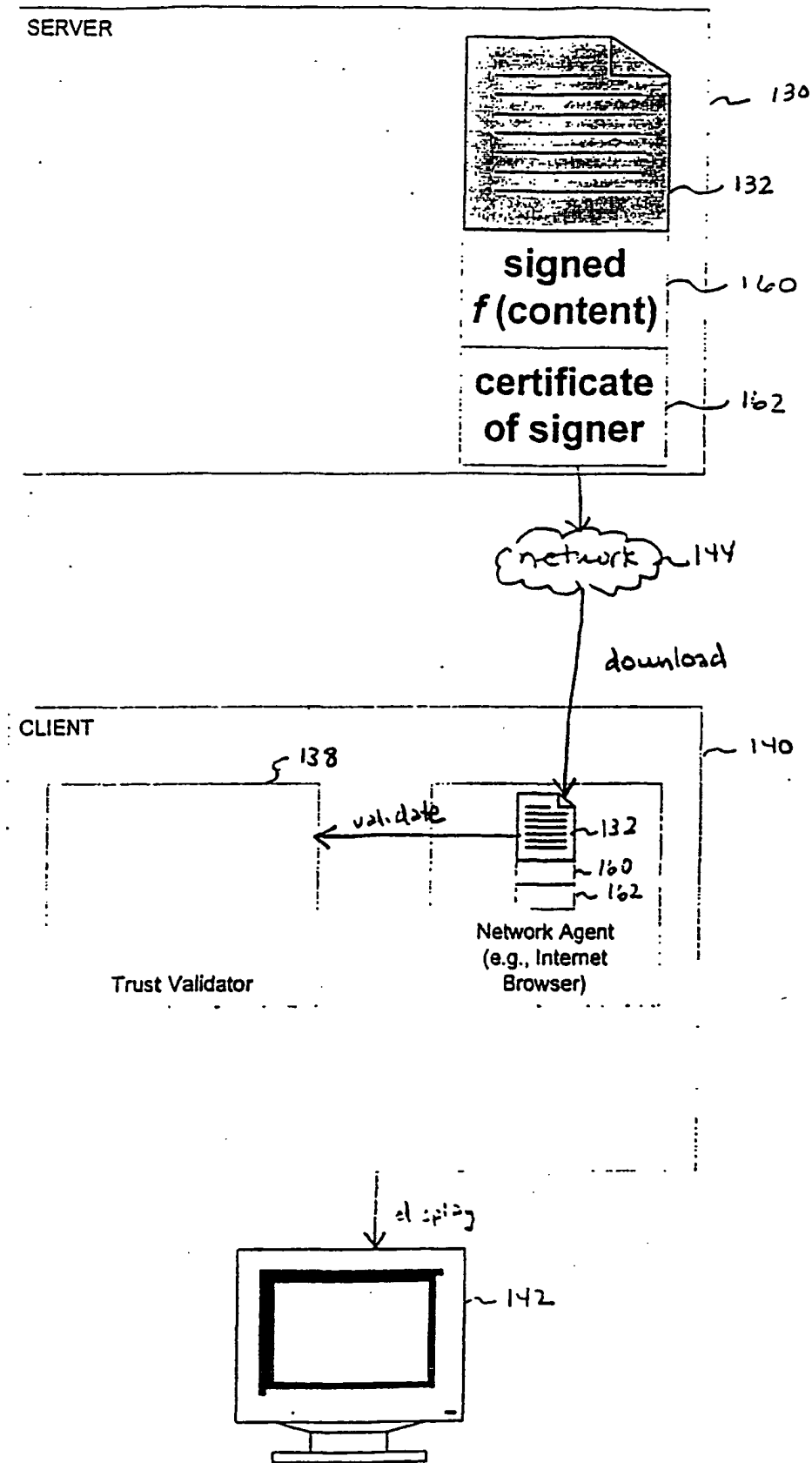


FIG. 11

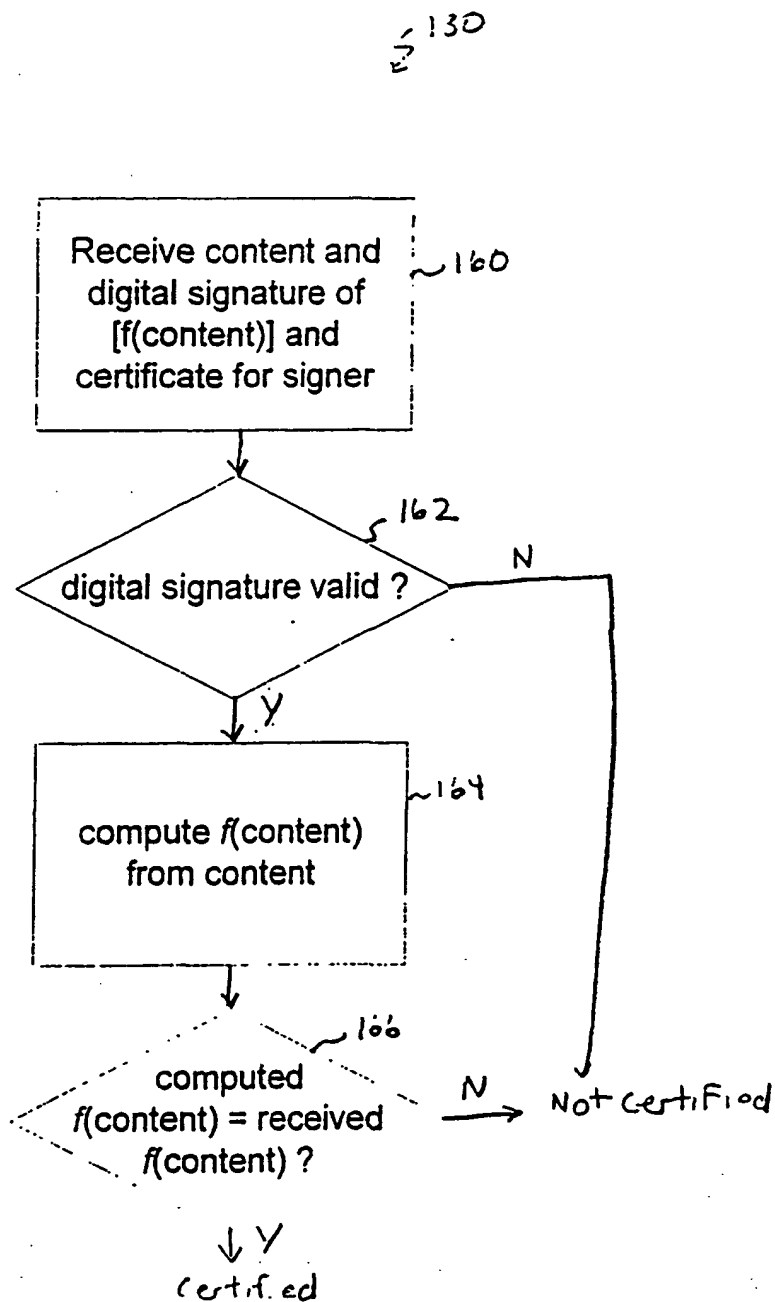


FIG. 12

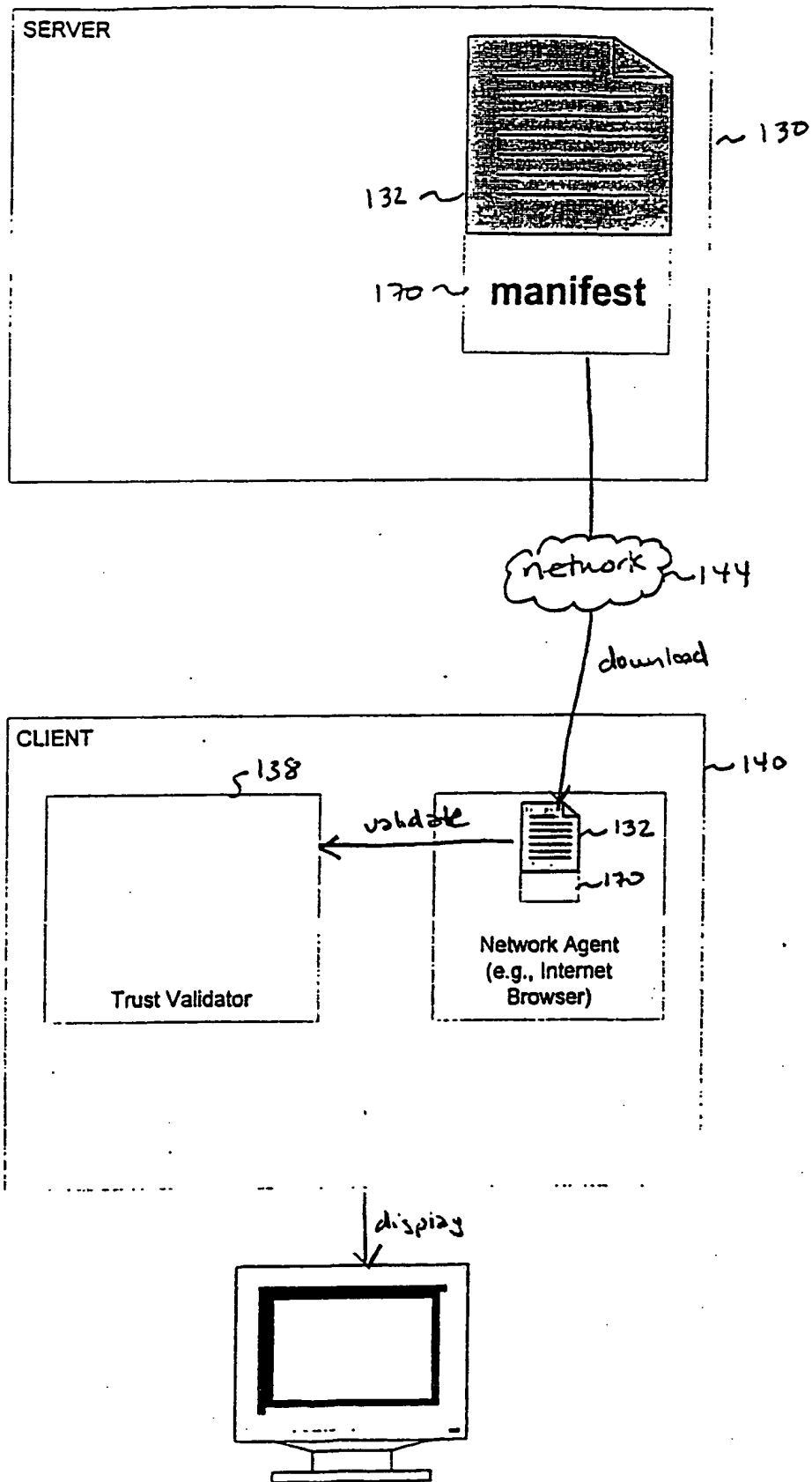


FIG. 13

5 170

Contents	f (content)
www.fr.com	984emf9
www.fr.com/digests.gif	29482jd9
www.fr.com/publications.gif	2930843f
www.fr.com/about.gif	23901233
.	
.	

FIG. 14

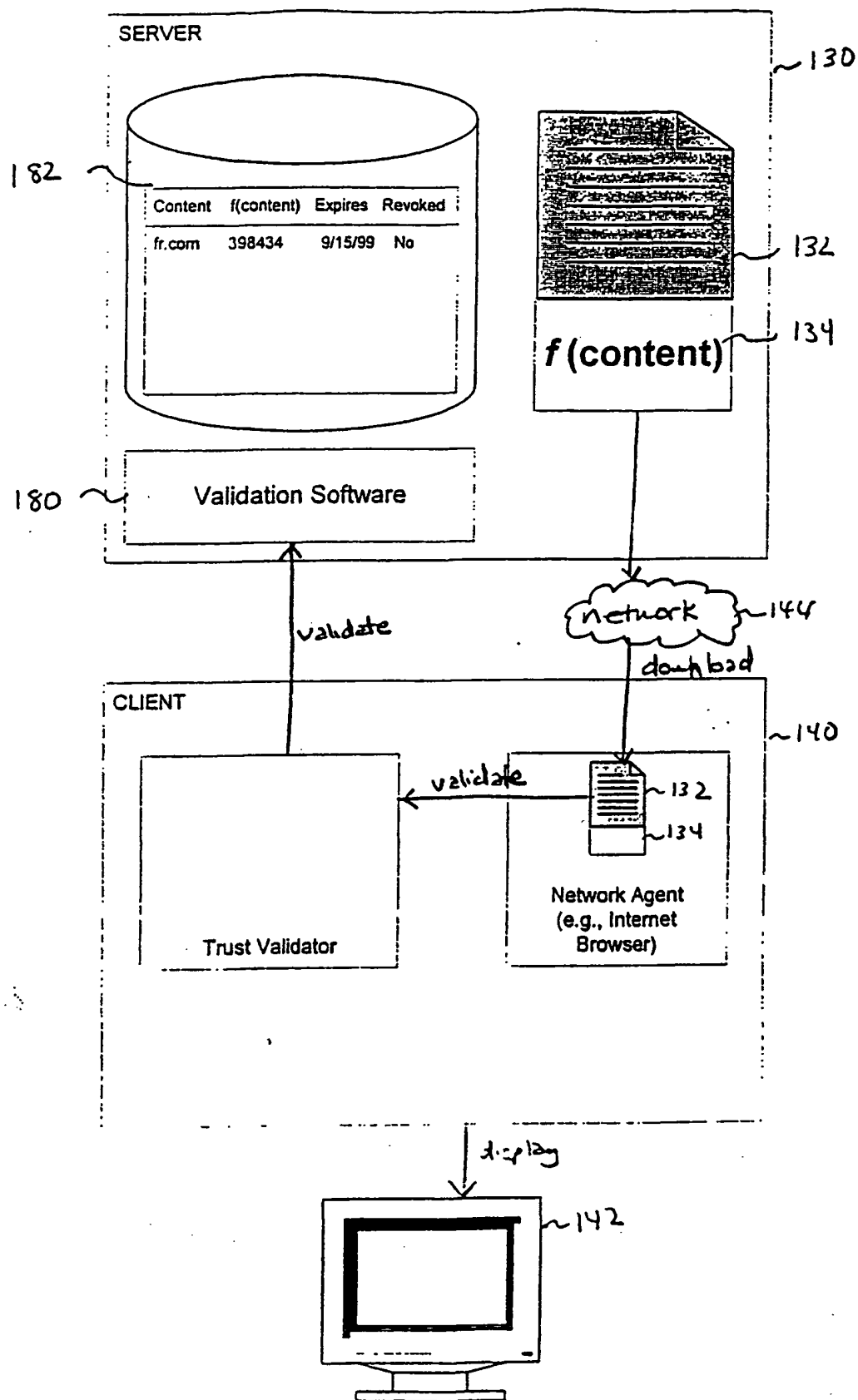


FIG. 15

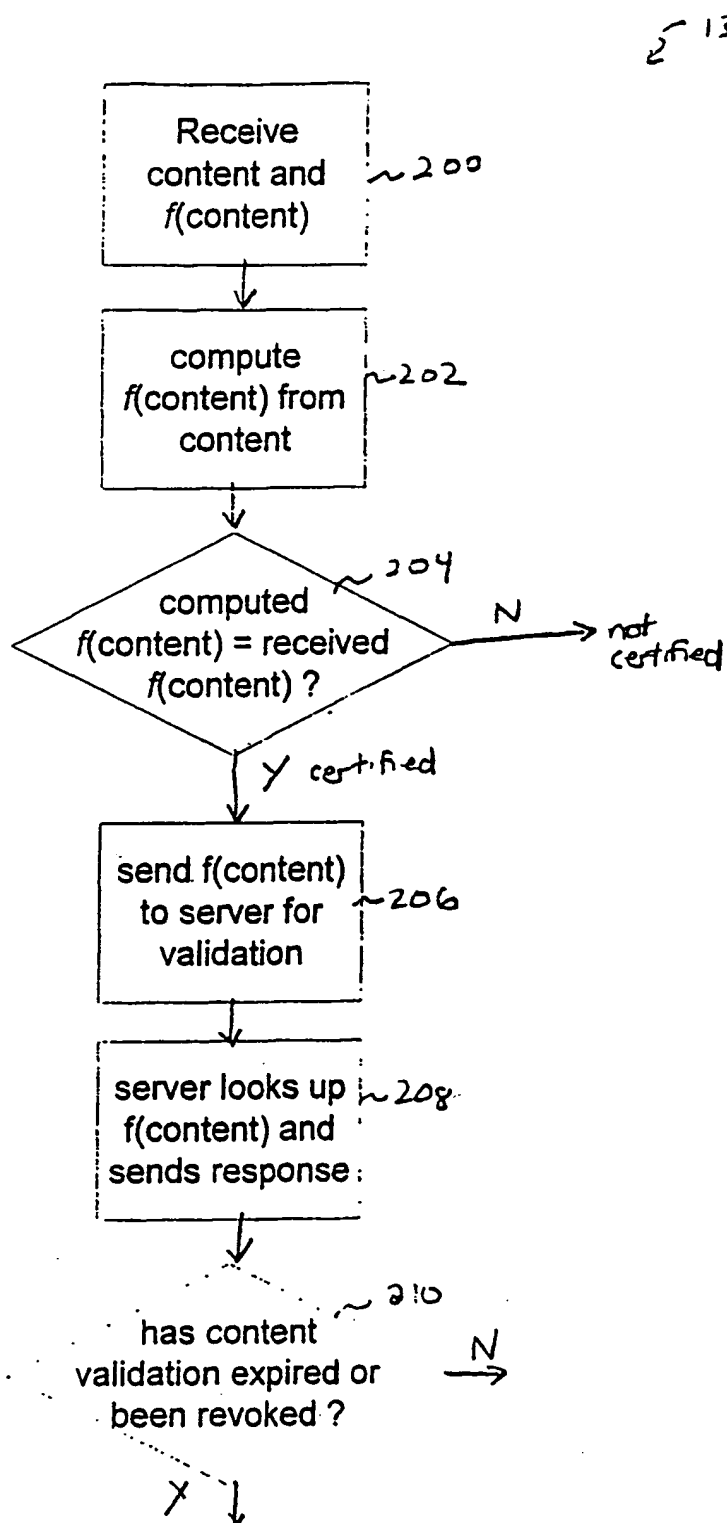


FIG. 16

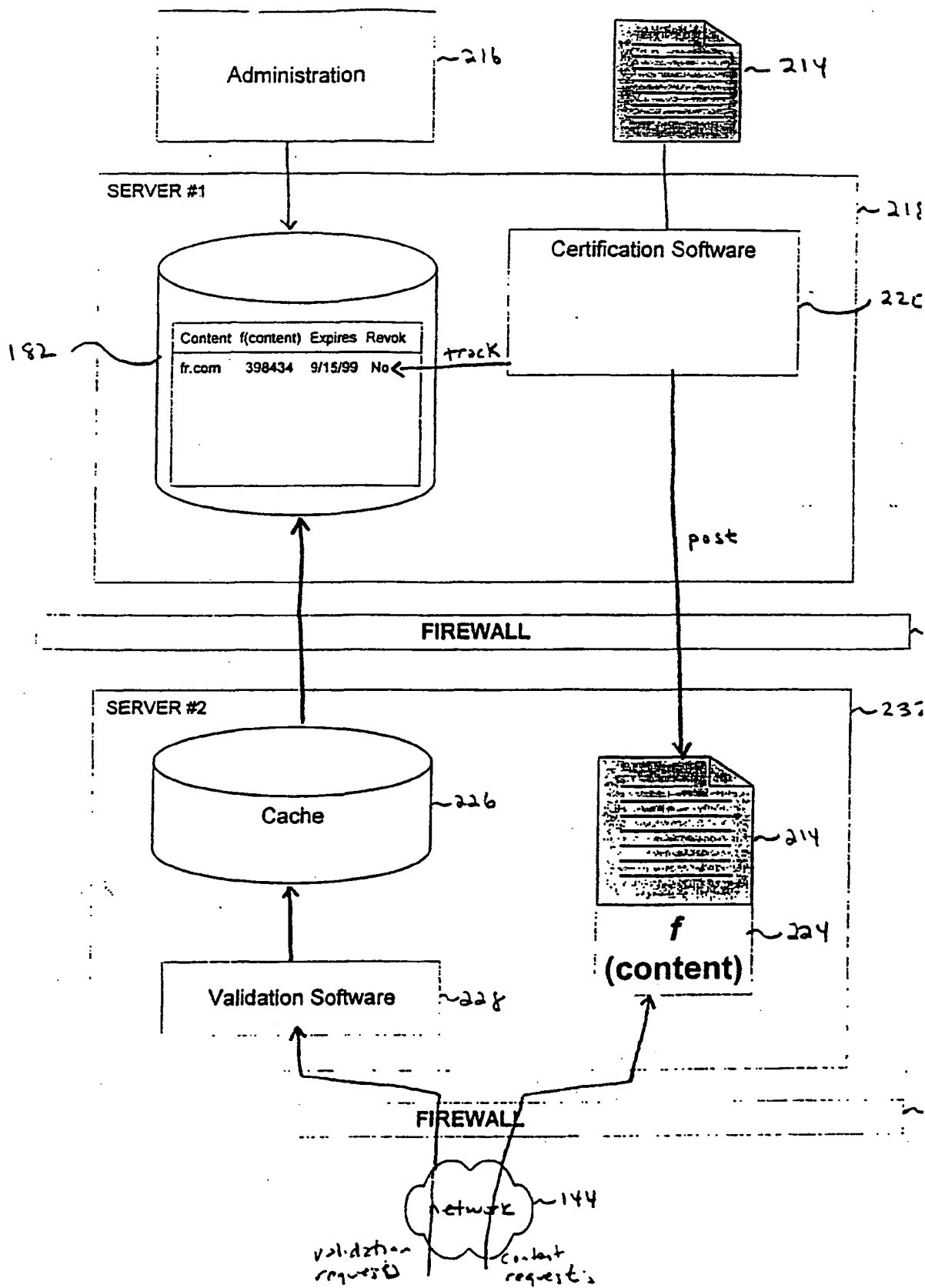


FIG. 17

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/10886

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : H04L 9/00, H04K 1/00, G06F 19/00

US CL : 705/44, 380/25

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 705/44, 380/25

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6,058,383 A (NARASIMHALU et al) 02 May 2000, fig. 1, col. 5, lines 11-40 and col. 6, lines 32-46).	1-6
Y	US 6,026,166 A (LEBOURGEOIS) 15 February 2000, col. 3, lines 46-col. 12, line 65.	1-6
Y	US 6,018,724 A (ARENT) 25 January 2000, col. 2, line 42- col. 22, line 22.	1-6



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

A document defining the general state of the art which is not considered to be of particular relevance

B earlier document published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

Z

document member of the same patent family

Date of the actual completion of the international search

21 MAY 2001

Date of mailing of the international search report

02 AUG 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

GLENTON BURGESS

Telephone No. (703) -305-4792

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.